图书情报工作
ChinaXiv合作期刊
第 65 卷 第 23 期　2021 年 12 月

数字背景下《春秋经传引得》错漏校订及编纂质量分析*

■ 彭秋茹[1,2]　范文洁[1,2]　黄水清[1,2]

[1] 南京农业大学信息管理学院　南京 210095　[2] 南京农业大学人文与社会计算研究中心　南京 210095

摘　要：[目的/意义]在数字化的前提下对《春秋经传引得》的错漏进行识别和校订，以量化数据为依据整体考察和分析其编纂质量，并得以管窥手工时代中国古代典籍索引扛鼎之作《汉学引得丛刊》的编纂质量，重新发现印本形态的古籍全文索引在数字时代的价值。[方法/过程]对《春秋经传引得》的经文、传文以及全部索引款目全面数字化，并逐项比对，找出并校订其中的错漏，记录并统计错漏的类型、数量，分析整体编纂质量。[结果/结论]《春秋经传引得》确有极少的错漏，款目差错率仅为万分之一左右，编纂质量堪称手工时代的巅峰，当得起学界的高度赞誉，可作为数字化语料的高质量基础语料。

关键词：《春秋经传引得》　校订　编纂质量　数字化　人工语料
分类号：G256
DOI：10.13266/j.issn.0252-3116.2021.23.006

1　引得及其编纂质量

"引得"即索引，上世纪初由时任燕京大学图书馆主任洪业先生根据英文"index"一词音译而来。上世纪 20 年代，受"西学东渐"之风的影响，特别是在"五四"新文化思潮的强烈冲击下，引用西方索引法编制、利用和研究索引在中国学术界渐成共识。在此背景下，洪业先生 1930 年创办了哈佛燕京学社引得编纂处，主持编纂了中国古代典籍的索引丛书——《汉学引得丛刊》，从 1931 年至 1951 年共出版了正刊 23 种、50 册，特刊 41 种、31 册。《汉学引得丛刊》的索引类型包括逐字索引、重要词汇索引、引书索引以及书名、篇名、类名、人名、地名索引等[1]。由于《汉学引得丛刊》在学界享有盛誉，"引得"有时专指《汉学引得丛刊》。

1.1　《汉学引得丛刊》

《汉学引得丛刊》是众多古籍索引中的一种。古籍索引将古籍中的有关事物名称、篇名、字句、词语、人名、地名、内容主题等分别摘录标引，注明出处、页码与行数，并按一定的排检方法编辑而成，以供人们查询有关古籍文献的内容[2]。古籍索引是揭示古籍内容的一种方式，在此之前也称为"备检""通检"。我国最早的古籍索引可以追溯到明代万历年间张士佩编制的《洪武正韵玉键》[3]，该书将《洪武正韵》的字头按照部首重新编排生成部首索引表[4]。明末傅山的《两汉书姓名韵》是中国现存最早的人名索引，清代蔡烈先的《本草万方针线》是我国现存最早的中医古籍索引。章学诚专门为其著作《历代纪年经纬》编制了专书索引《历代纪元韵览》，清代汪辉祖的《史姓韵编》是专门为二十四史中的人物传记所编制的索引。在《汉学引得丛刊》之后，中法汉学研究所通检组编纂了《论衡通检》《吕氏春秋通检》等十余种通检，顾颉刚编纂了《尚书通检》，商务印书馆出版了《十通索引》《佩文韵府索引》等。

在现有的古籍索引中，《汉学引得丛刊》不但体量最大，内容最全，而且具有选本精、款目细、内容全、使用方便、著录灵活、学术性强、除专书引得外还有专题引得等特点[5]。《汉学引得丛刊》开创了运用现代科学方法编纂中国古籍索引的新途径，改变了古籍索引编纂分散、个别、缺乏系统性的缺点，将编纂工作纳入规范、合理的科学轨道，并在出版形式、体例、编排设计、索引类型等方面作了大量创新，取得了令人瞩目的成果[6]。美籍华人陈毓贤女士在《洪业传》中写道："如果我们说这些引得在中国研究古籍的学术上创立了新纪元，可能也不算夸大其词。"[7]

* 本文系国家社会科学基金重大项目"基于《汉学引得丛刊》的典籍知识库构建及人文计算研究"（项目编号：15ZDB127）和国家社会科学基金青年项目"中文人文社会科学学术图书全文本引文分析模式研究"（项目编号：20CTQ026）研究成果之一。
作者简介：彭秋茹，助理研究员；范文洁，博士研究生，通讯作者，E-mail：2018214003@njau.edu.cn；黄水清，教授，博士生导师。
收稿日期：2021-08-17　修回日期：2021-09-29　本文起止页码：50-57　本文责任编辑：易飞

彭秋茹，范文洁，黄水清. 数字背景下《春秋经传引得》错漏校订及编纂质量分析[J]. 图书情报工作,2021,65(23):50－57.

《汉学引得丛刊》既植根于中国索引学的传统又受西方现代索引科学的影响,涵盖了周遍型字词引得、综合性关键词引得、书名(篇名)引得、图谱引得、人物姓名与字号引得等类型多样的引得体系,在学术史上具有重要地位。无论是在引得的类型、引得的款目,还是引得的排检法等方面,《汉学引得丛刊》都别具匠心。尤其是《汉学引得丛刊》对古典文献中的人名、地名、机构、时间及其他具有语义的词全面、系统地进行了索引,其庞大的体系、一流的质量令后人至今仍难企及。

《汉学引得丛刊》问世以后,倍受学界推崇,学者纷纷不吝溢美之辞。无论是对《汉学引得丛刊》的系统评价[8],还是使用价值评判[9],或者评述《汉学引得丛刊》的体系、功能、特点、学术地位[5,10－12],后人的文字满满的都是赞誉。对于《汉学引得丛刊》的不足,学者们普遍诟病的是洪业先生独创的"中国字庋撷"检字法,认为其较为繁琐,事实上该检字法并未成功推广。另外,也有学者认为部分引得收书存在不妥,每部引得的卷首序文太长且与引得本身无关,应该单独刊行[6]。但无人质疑过《汉学引得丛刊》的编纂质量。

1.2 编纂质量

古籍索引的一切价值和地位都以编纂质量可靠可信为前提。人工编制古籍索引不但是一项技术性很强的工作,同时也是一项耗费心力且须投入大量时间的工作,长期的细致寻检和繁琐统计,既须聚精会神又枯燥乏味[13],即便如此,编纂质量仍难以保证。《汉学引得丛刊》的编纂质量却赢得了学术界的普遍赞誉。这是因为洪业先生为《汉学引得丛刊》制定了完整、严格的编纂方案,共分为选书、选本、标点、抄片、校片、编号、稿本、印刷、印本校对、加序 10 个步骤[14],形成了行之有效的编纂制度,用于具体指导引得的大规模编纂工作。需要指出的是,《汉学引得丛刊》由于是人工编纂,一致性难于保证,从理论上推测,出现错漏在所难免,甚至还应该比较常见。

学界一般认为,64 种 81 册《汉学引得丛刊》的编纂质量虽非至善至美,但错漏之处已经很少[6],然而学者的这些判断均为个人主观感受,并没有充分的数据支持,目前暂未发现有人以量化数据整体考察某种或某册《汉学引得丛刊》的编纂质量、分析可能的条目错漏。事实上,整体考察与分析引得编纂质量需要比对原始文献和所有条目,劳心费力,若以人工方式实现,其工作量几乎等同于重新编纂一部引得,在手工时代若试图开展这方面的研究存在巨大的困难。

技术的进步使得开展此项工作成为可能。数字人文时代,数字化的原始文献较易获得,《汉学引得丛刊》的重要性也重新凸现,各项技术手段也足于支撑索引款目的录入、生成、校对、比较、分析、统计等工作,整体考察与分析《汉学引得丛刊》的编纂质量既有现实意义,也具备技术可行性。

20 世纪 80 年代中后期,由于全文检索技术得到广泛应用,书本式手工检索工具一度受到冷落,《汉学引得丛刊》为代表的古籍索引也一定程度上被学界忽视。但是,全文检索技术大幅度提高查全率的同时,在查准率方面却难于令人满意,而《汉学引得丛刊》以人名、地名、机构、时间序列及其他具有语义的词汇为索引词全面揭示了文本中蕴涵的语义知识,不仅可以提高查准率以弥补全文检索技术的不足,在数字时代还可以被看作高质量的古代汉语文献人工语料库,可将其数字化成为开展典籍文本人文计算的基础语料。人文计算的基础语料比手工检索工具对质量要求更高,不允许将印本时代的错误带入数字时代,因此,整体考察与分析《汉学引得丛刊》的编纂质量,校订其中的错漏非常必要。

本文选择《汉学引得丛刊》中的《春秋经传引得》为对象,在对其全面数字化的基础上获得相关数据,形成数字形态的结构化人工语料,以识别并校订其中的错漏,整体考察与分析《春秋经传引得》的编纂质量,并籍此管窥《汉学引得丛刊》的总体编纂质量。

2 《春秋经传引得》的数字化

《春秋经传引得》是《汉学引得丛刊》中最重要、最具价值的系列——"十二经引得"(缺十三经中的《尚书》,因为洪业认为《尚书》是伪书)中的一种。选择以《春秋经传引得》来测评《汉学引得丛刊》的编纂质量有典型性和代表意义。《春秋经传引得》原书于 1937 年由哈佛燕京学社初刊,1983 年上海古籍出版社以影印方式再版。

《春秋经传引得》属于《汉学引得丛刊》中的特刊,原书包含了《春秋》经以及公羊、谷梁、左传三传的全部文本。经、传原文汇集在书前的"春秋经传"部分。《春秋》经文以及三传的传文根据鲁国的十二公分为十二篇,每篇又根据每个公的在位年数分为若干章,每章又根据具体内容分为若干节,每节前面会用一个阿拉伯数字作为一个标识。

《春秋经传引得》将《春秋》经、传中的句子逐字、逐词设立引得款目,编成综合性索引,每条款目详注

字、词所在的文句及页码、各公年代和节数。以下是《春秋经传引得》的引得款目的实例(见原书第165页):

子蒲

秦子蒲○○子虎帅车五百乘以救楚,446/定5/附2

○○曰,446/定5/附2

子期○○滅唐,447/定5/附2

其中,“子蒲”为标目,在“子蒲”这个标目下共有3个句子,为经文与传文中所有包含“子蒲”一词的句子,可称之为“条目”。标目与条目共同构成引得(索引)款目。条目中的“○○”用于指代标目“子蒲”。条目逗号后面的数字和文字标识了条目在经文和传文中的原始出处。

2.1 数字化的目的与思路

为整体考察与分析《春秋经传引得》的编纂质量,首先要将《春秋经传引得》的经文、传文以及所有引得款目全部数字化,以便逐项比对,找出错漏。数字化过程共可细分为3个阶段,即经、传原文的数字化、引得标目的数字化、引得款目的自动生成及校对。

首先在网上寻找并下载与引得编制时选择的经、传原始版本相同或者相近的古汉语文本,再与《春秋经传引得》经文和传文逐字比较、校对,还原引得编纂时使用的原始文本。这个阶段要解决引得原文排检错误、异体字、原文页码添加、句读信息等问题,以确保第3阶段生成的每一条索引款目与原始引得具有可比性。

其次,人工录入索引的标目,生成索引标目文档。

最后,通过计算机程序,由标目文档生成初始引得款目文件,再将初始引得款目文件与《春秋经传引得》的各条款目一一比较、校对,记录编纂时存在的错误,形成最终的引得款目文件。

2.2 经传原文的数字化

《春秋经传引得》的经、传原文依据的是1926年上海锦章书局所印《十三经注疏》(附《校勘记》),该版本为嘉庆二十年(1815)阮元南昌府学刻本《十三经注疏并附校勘记》的影印本。

为得到《春秋》经、传原文的数字化版本,首先通过网络爬虫从“中国哲学书电子化计划”网站(http://ctext.org/)下载了《春秋》经及三传的繁体字版本的正文,再按照《春秋经传引得》中“春秋经传”的篇、章、节的顺序对经、传的文字进行整合。在此基础上,依据上海古籍出版社1983年重印的《春秋经传引得》的正文人工进行格式和文本校对[15]。

“中国哲学书电子化计划”网站上的《春秋》经及三传皆本武英殿版《十三经注疏》,与《春秋经传引得》原文依据的阮元南昌府学刻本并不一致。以“中国哲学书电子化计划”的《春秋》经及三传原文作为数字化的初始文本,可以避免从零开始的手工录入,把录入工作简化为校对工作,减少了数字化过程的工作量。校对过程以《春秋经传引得》原文为依据,用《春秋经传引得》原文去校正从“中国哲学书电子化计划”网站下载的数字文本。也就是说,经校对后得到的是与《春秋经传引得》阮元南昌府学刻本完全一致的《春秋》经及三传的数字化文本。

《春秋经传引得》的经、传原文以经文为纲领,注明三传与经文的异同,然后按照《公羊》《谷梁》《左传》的顺序依次排比列于经文之下。全部经、传原文总共九万七千余字。经、传文句若有异文见于阮元的《校勘记》的,以脚注的方式注明异文见于《校勘记》的卷数、页码。

经、传原文在《春秋经传引得》的编排形式如下:

--*公--

{*公*年}

1 元年/春,王正月。(经)

公元年者何? 君之……(公羊传)

穀雖無事必舉正月,謹始也。公……(穀梁传)

左元年,春,王周正月,不……(左传)

为了在计算机自动生成索引款目时确保句子的完整性和准确性,对下载的经、传原文的句读还采用人工校对的方式作了如下处理:若下载的经、传原文与《春秋经传引得》正文的断句一致,则保留标点符号;若断句不一致,则根据实际情况或者添加断句标识“/”或者直接删除原标点。《春秋经传引得》正文中用于索引款目断句的符号有:“,”“。”“:”“?”“/(校对过程中加的断句标识)”,不用于索引款目断句的符号有:“、”“「”“」”“《”“》”“『”“』”“<”“>”。

2.3 标目的数字化

标目是索引款目中用于排检的项目,也是手工时代款目的检索入口点,洪业先生在《汉学引得丛刊》中称之为“目”,也即引得款目第一行的首字词。为使计算机能自动生成《春秋经传引得》引得款目,首先需要将《春秋经传引得》全部标目逐字、逐词手工录入计算机,形成相应的文档。

在《春秋经传引得》的原始款目中,当标目存在一词多义现象时原则上是不作区分的,但如果这个标目

为特殊名词,则需要在相应的标目后面添加对该词汇的特殊释义,并将特殊释义用括号括起,单独设立标目,以区别于常用词义。在人工输入标目时,完全遵循了这一处理方式,其中的括号在输入时采用全角中文字符。例如,“心”在《春秋经传》中有两种释义:一是与现代汉语的常用词意思相同,作“心肺”“心意”等解释;还有一种则是“星宿名”,即二十八宿的“心宿”,如《左传》襄公九年的“或食于心”及“心为大火”。在《春秋经传引得》中,这两种情况分别设立标目,后一种情况的标目词“心”后加括号,注明“星名”。

此外,《春秋经传引得》的某些标目后还存在括号加“参”的表示方式,目的是为了标识和处理人名的异名同指(一人数称)、同名异指(数人同称)现象。

异名同指,即一人数称,也即同一人有多个称谓。这种情况下有正名的以正名为主,无正名的则取其字或者号为主,王侯有谥号的以谥号为主。所有异名或异称都单独设立标目,后面用括号加“參”指向其他称谓。主要称谓标目的“參”后面标注所有非主要称谓,非主要称谓标目“參”后面只标注主要称谓。

例如:

宋襄公(參:茲父,襄公,宋子,宋公,宋公慈父)

　宋襄公卽位

宋襄公問焉

……

茲父(參:宋襄公)

茲父之不葬何也

襄公(參:宋襄公)

　未葬而襄公會諸侯

　與襄公之征齊也

宋子(參:宋襄公)

　公會宰周公齊侯宋子衛侯鄭伯許男曹伯于葵丘

　以宋子爲無哀矣

　……

宋公(參:宋襄公)

公會齊侯宋公陳侯衛侯鄭伯許男曹伯于鹹

公會齊侯宋公陳侯鄭伯許男曹伯盟于牡丘

……

宋公慈父(參:宋襄公)

　宋公慈父卒

宋襄公在经传原文中共有“宋襄公”“兹父”“襄公”“宋子”“宋公”“宋公慈父”6种称谓。“宋襄公”是其谥号,而“兹父”“襄公”“宋子”“宋公”“宋公慈父”为别称。原文中无论出现哪种称谓都单独设立标目,当别称作为标目时用括号加“參”指向正名“宋襄公”,正名“宋襄公”标目中则用括号加“參”列举出所有5个别称。

同名异指,即数人同称,也即多人共用一个称谓。这种情况须在括号中标注国别以作区分。大多数情况下,同名人的还会有其他的称谓,须采用异名同指的处理方式,以在国别后面加“參”的方式标注,以示区别。

例如:

文伯(晉,參:士文伯)

　文伯復命(襄公三十一年)

文伯(晉,參:荀躒)

　以文伯宴(昭公十五年)

文伯揖籍談(昭公十五年)

文伯(魯,參:穀)

　生文伯(文公七年)

　魯人立文伯(文公十四年)

……

“文伯”一名均出自《左传》,但在《左传》原文中不同的场景所指代的人物不同,共3个不同的“文伯”。前两个“文伯”分别是晋国的“士文伯”和“荀躒”,后一个“文伯”指“魯”国的“穀”。每个标目词“文伯”之后的括号中分别标明国别“晉”和“魯”,再在国别后加“參”表明该名称分别指向“士文伯”“荀躒”和“穀”。

异名同指和同名异指的标目,也完全按照《春秋经传引得》原始标目的样式录入,括号及其他标点符号也均采用全角中文字符。

2.4 款目的自动生成与校对

编写计算机程序,依次自动读入标目文档的各条标目,再将标目与数字化的经、传原文文档相匹配,自动生成并输出对应的引得款目。

由于计算机程序只能通过字形来匹配相应的款目信息,而无法区分同形异义的情况,且程序自动生成的款目可能存在错误匹配,需要通过人工方式对每个标目下的每条款目进行校对,确保每条款目完全正确,如此,得到完整的引得款目文件。

在校对款目时使用“互参验证”的方法将标目、《春秋经传引得》、数字化的经传原文以及自动匹配生成的款目结合起来进行校对。具体实施步骤见图1。

整个校对流程从自动生成的款目开始,首先与纸本版的《春秋经传引得》进行对比,检查每一条款目信息是否一致。如果一致,则标目下的款目视为完备准确;若不一致,根据索引款目的组成,那么可能存在4个方面的问题:①标目错误。因为标目最初是由手工

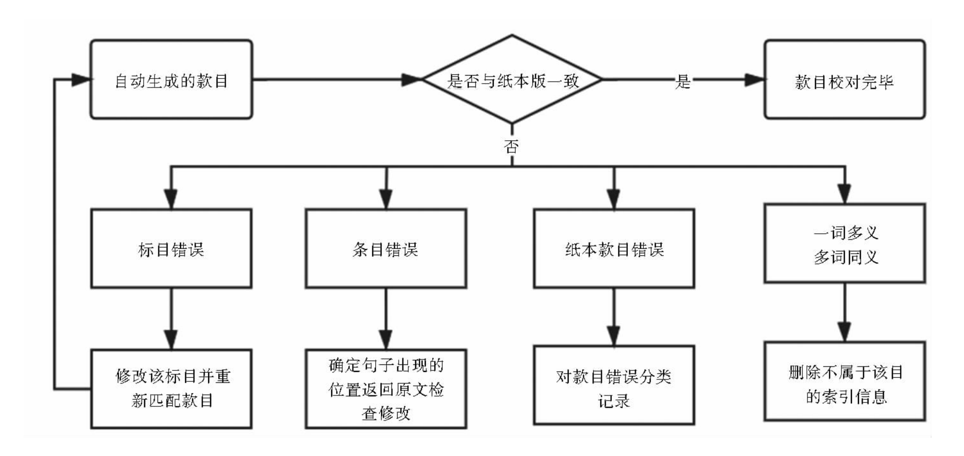


图 1　互参验证的款目校对流程

输入的，并且 1983 年重印的《春秋经传引得》由于印刷质量问题部分字迹模糊，难以确定字形，可能出现输入错误。因此在款目校对时，再次检查该标目，并重新自动匹配生成款目，实现二次校对。②条目错误。即该标目下的条目个别字词与原文不一致，需要返回经、传原文重新修改，并对该款目进行修改。③纸本款目错误。即纸本《春秋经传引得》本身出现错误。在排除了标目和条目错误后，如果仍出现款目不一致，则是《春秋经传引得》编纂过程中出现了错标、漏标、错抄、错排等。这些错误在当时的条件下是很正常的现象。据《春秋经传引得・序》记载："全文校印即就，乃逐句剪贴片上。一字为句者，得一片；数字为句者，片如其字之数；合得二十余万片。"可见引得编纂过程细节繁复，若出现剪贴片的遗失，则很容易造成索引款目遗漏。并且，剪贴片是由多人同时抄录的，因而同一字词可能会由于抄录者不同而出现字形不一致的情况。④计算机自动识别带来的问题。计算机只能通过文字编码来识别汉字，但是无法判断其具体语义，因此，对于一词多义以及多词同义的标目，还需人工去核对其下的款目是否完全正确。

经过经传原文的数字化、索引标目的人工录入以及索引款目的自动生成和校对，最终得到以数字化形态呈现的完整、正确的《春秋经传引得》，共包含索引标目 13 764 个，引得款目 180 162 条，总字数达 410 万。对其中的异体字以及与经、传原文用字不同的情况也做了相应的标注。

3　《春秋经传引得》的错漏及校订

在《春秋经传引得》的数字化过程中，通过互相参照验证校对款目，可以发现并记录《春秋经传引得》中出现的各种错漏。从错漏的类型看，推断主要是编纂过程中不同编者的用字习惯、编者之间缺乏沟通、编纂技术不成熟、缺乏相应的校验过程以及同一标目下的款目数目繁多等多种原因造成的。以下按错漏的类型分别列举：

3.1　并用异体字

并用异体字是指一个字有多种字形，而在《春秋经传引得》中不同的场所中采用了不同的字形，使得在检索的时候造成漏检。在《春秋经传引得》中，存在并用现象的异体字共有 5 对，它们分别是"甯"与"寗"、"恒"与"恆"、"嘗"与"甞"、"啓"与"啟"、"閒"与"間"。

"甯"作为姓氏的时候还可以写作"寗"。"甯"在标目中作为首字和非首字共出现了 30 次，其中，姓氏 17 次，人名 3 次，地名 1 次，参照项 8 次，普通释义 1 次。在经、传的原文中总共有 79 个句子包含"甯"（"寗"），其中"甯"8 句，其余为"寗"，但在引得款目中全部用的是"甯"。

"恒"（"恆"）在标目中共出现了 4 次，分别是"恒""恒(參:陳恒)""恒事""恒星"，统一用"恒"。但在经、传原文中，"恒"与"恆"两个字是并用的，总共有 20 个句子含有这两个字，其中 4 句用"恆"，剩下 16 句用"恒"。

"嘗"与"甞"简体写法均为"尝"，在《春秋经传引得》的索引部分只有"嘗"，但在经、传原文中"嘗"与"甞"两种字形都有，其中"嘗"出现了 35 次，"甞"出现了 25 次。除去"桓公十四年"这一章的"甞"有 10 处注有"甞一作嘗"外，表示"甞"也可能写作"嘗"，其他章节出现的"甞"均未见有注。

"啓"与"啟"在经文和传文中总共出现了 71 次，其中"啟"出现 25 次，"啓"出现了 46 次。作为标目的首字时原书中出现的都是"啟"，如"啟""啟(宋)""啟

(宋,華氏)”“啟(楚,參:公子啟)”“啟服”,但在诸多不是首字的人名标目中原书用了“啓”,如“微子啓”“薳啓彊”“北郭啓”。同时,在对应的经传原文中还有“薳啓彊”与“薳啟彊”两种写法。可见编者在编纂引得的时候并没有区分两者,而是随机并用两种字形。

汉字在简化前后也存在一些差异,比如“閒”与“間”。“間”字出现得很晚,凡是后来写作“間”的在先秦文献中都应写作“閒”,最早不超过乾隆年间才将 jiān 和 jiàn 写作“間”,把 xián 写作“閒”。显然在编纂引得的时候编者并没有考虑到两者的区别,在原书中两者均有出现。“間”共出现 6 次,其中,僖公三十年有“吾子取其麋鹿以閒敝邑”,该句下注有“閒一作閑”,其余“閒”均表示“間”之意,但是在索引部分只见“閒”而不见“間”。

异体字并用造成了自动生成索引时的款目不全和遗漏,需要通过人工的翻阅引得才能发现端倪。究其原因,可能有三:①上述几对异体字在当时用法上没有差异,导致编者没有意识到字形差异造成的翻检不便;②在索引编纂时由于分工不同,没有做到编者之间协调统一;③受索引编纂技术的局限,索引编纂完成之后没有快速有效的方法查验索引编纂过程中存在的错误和偏差。

3.2 款目与经传不一致

《春秋经传引得》中还存在款目与经传原文前后不一致的现象,全部索引款目中总共发现 10 处。与经、传原文的句子相比,前后不一致表现为索引款目存在缺字、错字、字序不一致。如表 1 所示:

表 1　索引款目与经传原文不一致

类型	原文	款目(括号内为标目)
缺字	楚子入居于申	子入居于申(入)
	小人務知小者近者	小人知小者近者(小)
错字	未圍齊也	未爲齊也(爲)
	仲行伯之於晉也	中行伯之於晉也(中行伯)
	請以死告	詰以死告(死)
	琴張聞宗魯死	秦張聞宗魯死(死)
	大有巡功	夫有巡功(巡)
	文公子佗殺太子免而伐之	文公子佗殺太子免而代之
字序不一致	君將以文之舒鼎	將君以文之舒鼎(之)
	弟兄三人	兄弟三人(人)

缺字是指在索引句与较原文缺少字词,导致语句不完整。例如,在标目“入”下有“子入居于申”,原文位置为“第 131 页,僖公二十八年,左传”,在经传中对应的原文为“楚子入居于申”,索引款目中缺少了“楚”字。错字不同于并用异体字,并用的异体字两种写法均是正确的,仅仅是字形不一致,而错字则是在索引中错用了别的汉字。错字的情况共发现了 6 处,错误的发生原因为同音或形似。如“圍”与“爲”同音,“仲”与“中”同音,“琴”与“秦”同音,但是它们既不属于通假又不属于异体。而“請”与“詰”字形似,“大”与“夫”形似,“伐”与“代”形似。最后还有一类是经传原文与索引款目中的字词顺序出现不一致,这类错误共有 2 处。在“之”标目下将原文“君將以文之舒鼎”误写为“將君以文之舒鼎”,在“人”标目下将文公十一年《谷梁传》中的“弟兄三人”误为“兄弟三人”。

这类错误是编者手工编纂时的疏忽,编纂完成后也无法逐一校对。但是这类错误相较于十万级的索引款目已经是微乎其微了。

3.3 遗漏条目

遗漏条目是指在同一个标目下缺失了某条或某几条本该有的条目。在《春秋经传引得》中只有标目“之”下面有这种情况。这一方面可能是由于编者的疏忽,另一方面“之”是索引款目最多的标目,共有 8 261条之多,且一般作代词和助词,故可能出现遗漏。

具体地说,共遗漏和缺失以下 3 条条目:哀公七年,“獲白鳫,獻之,且言田弋之說,說之”,遗漏和缺失了条目“說之”;哀公八年,“若有之,則分室以行”,遗漏和缺失了条目“若有之”;哀公八年,“及潞,麇之以入,遂殺之”,遗漏和缺失了条目“遂殺之”。

3.4 句读前后不一

洪业先生在《春秋经传引得・序》中表示全书的句读都经过校订,但是经过数字化以后的校对仍然发现了 4 处句读前后不一的情况。有 2 处是对经、传原文切分了的地方引得款目中未作切分,对语义的理解也不产生歧义。其一原文为“胡簋之事。則嘗學之矣。甲兵之事。未之聞也。”但在“之”字标目下该句被断句为“胡簋之事則嘗學之矣。甲兵之事未之聞也。”其二为宣公四年的《谷梁传》,有“不肯者。可以肯也。”而“也”字标目下该句为“不肯者可以肯也”。还有一处是索引款目断句错误。《左传》昭公二十年有“宋元公無信多私而惡,華向、華定、華亥與向寧謀”,而“向(参:向氏)”标目下的该句被断为“宋元公無信多私而惡華向”,“華向”被切到了上半句,产生了歧义。最后一处是《春秋经传引得》的原文断句错误。《左传》襄公三十一年有“雖獲其志不。能終也”,该句正确的句读应为“雖獲其志。不能終也。”在所有的索引款目中,该句的断句则是正确的。

4 结语

经过数字化后的详细的比对，发现了 410 万字的《春秋经传引得》全部 4 种类型错漏。其中，并用异体字略为常见，5 组异体字涉及 100 多个句子，但在引得款目中基本上都已统一用字，只有“啓”与“啟”在引得款目仍存在并用现象。除此以外的错漏还有：10 处款目缺字、错字和与经传原文字序不同，“之”字标目下遗漏 3 个条目，4 处经传原文与引得款目句读不一。

《春秋经传引得》卷帙浩繁，在手工时代，如此巨大的工作量，除去并用异体字现象外总共只出现了寥寥几项个位数的错漏，款目差错率仅为(10 + 3 + 4)/180 162≈0.009%，连万分之一都不到，令人瞠目。在经、传原文中并用异体字对于读者基本上不会产生太大的影响，只有引得款目中有异体字才会造成漏检。即使把出现在标目第 2、第 3 字位置的异体字“啓”(“微子啓”“蘧啓彊”“北郭啓”)的错误也包括进来，错误款目也仅增加了 13 条，款目的总差错率也仅从万分之一不到上升为(10 + 3 + 4 + 13)/180 162≈0.017%，即万分之二不到。

由于检索工具的特殊性，对于一致性的要求奇高，但《春秋经传引得》差不多做到了百分之百的一致。在没有计算机辅助的时代，全凭手、眼、脑达到如此高的准确性和一致性，洪业先生亲手制定的编纂制度功不可没。

虽然尚不能用一个具体的数值来评价古籍索引的编纂质量，但是通过上述对《春秋经传引得》所存在错漏的描述，不难得出结论：《春秋经传引得》极少错漏，其编纂质量创造了手工时代的巅峰，其质量和价值令后人望尘莫及！即使在计算机时代，要达到洪业先生当年的水准仍需要付出艰辛的努力。

本文选择《汉学引得丛刊》中的《春秋经传引得》为对象，通过录入、生成等过程，以数字化为手段整体考察与分析了《春秋经传引得》的编纂质量，同时也提出了一套规范化的古籍全文索引数字化方法，能够从庞大的古籍全文索引中发现和校订前人编纂时由于各种原因产生的错漏，并以数据为基础客观评价编纂质量。这样，既为读者和研究人员提供了使用方便的数字化人工语料，同时也保证了语料的准确性，从而使印本的古籍全文索引在数字时代价值得到充分发挥。以《汉学引得丛刊》为例，若全面数字化，不但能够校订错漏，满足读者快速、准确的检索需求，既减少误检又可以多角度呈现检索结果，更为重要的是，其庞大的周遍型语词索引及原文文本可作为语料库、知识库建设的基础语料，成为典籍文本人文计算的数据基础。如此，《汉学引得丛刊》中的语义知识将得到充分揭示，可助力中华优秀传统文化的普及与传播、社会主义文化建设以及相关学科的发展。

虽然《春秋经传引得》体量庞大，但只是 64 种、81 册《汉学引得丛刊》中的一种，篇幅也仅为 4 册。本文选择《春秋经传引得》为对象，研究编纂质量、校订错漏，进而以《春秋经传引得》为典型代表推定《汉学引得丛刊》的编纂质量，论据尚不够充分。后续研究应首先对《汉学引得丛刊》全面数字化，在此基础上实现对《汉学引得丛刊》编纂质量更全面、更客观的评价。

参考文献：

[1] 高娟，刘家真. 中国大陆地区古籍数字化问题及对策[J]. 中国图书馆学报，2013，39(4)：110 - 119.
[2] 黄建年，侯汉清. 基于 GB/T22466—2008 的古籍索引编制技术要点[J]. 图书馆建设，2011(6)：45 - 48.
[3] 陈东辉. 关于古籍索引工作的若干思考[J]. 北京图书馆馆刊，1997(1)：42 - 47.
[4] 赵生敏.《洪武正韵玉键》对明代后期字书部首法的影响[D]. 天津：天津师范大学，2019.
[5] 朱积孝. 哈佛燕京学社所编引得的内容与特点[J]. 江淮论坛，1985(3)：52 - 57.
[6] 顾志华. 试论洪业在索引学上的成就[J]. 文献季刊，2003(1)：232 - 243.
[7] 陈毓贤. 洪业传[M]. 北京：商务印书馆，2013.
[8] 王钟翰. 洪煨莲先生与引得编纂处[C]//学林漫步(第八集). 北京：中华书局，1983.
[9] 罗伟国. 发挥古籍作用必须编制索引[J]. 中国索引，2014(2)：4 - 5.
[10] 侯汉清，王雅戈. 中国近代索引研究的开山之作——《引得说》[J]. 大学图书馆学报，2006(5)：76 - 81.
[11] 邓德生. 哈佛燕京学社所编“引得”的得与失[J]. 贵图学刊，1988(4)：44 - 46.
[12] 马学良. 哈佛燕京学社汉学引得丛刊研究[D]. 保定：河北大学，2007.
[13] 陈东辉. 中国古籍索引工作的回顾与前瞻[J]. 辞书研究，1994(2)：83 - 91，126.
[14] 洪业. 引得说[M]. 北京：哈佛 - 燕京学社引得编纂处，1932.
[15] 洪业，聂崇岐，李书春，等. 春秋经传引得[M]. 上海：上海古籍出版社，1983.

作者贡献说明：

彭秋茹：论文框架及整体思路制定；
范文洁：数据整理与数据分析，撰写初稿；
黄水清：论文修改定稿。

Errors Revision and Compilation Quality Evaluation of *Index to Chunqiu Jingzhuan* in a Digital Background

Peng Qiuru[1,2] Fan Wenjie[1,2] Huang Shuiqing[1,2]
[1] College of Information Management, Nanjing Agricultural University, Nanjing 210095
[2] Research Center for Humanities and Social Computing, Nanjing Agricultural University, Nanjing 210095

Abstract: [**Purpose/significance**] Under the premise of digitization, this paper recognizes and revises missing errors of *Index to Chunqiu Jingzhuan*, whose compilation quality is reviewed and analyzed based on quantitative data. This paper shows the compilation quality of the great works of ancient Chinese classics index *Sinological Index Series* in the manual age and reveals the value of the printed form full-text index of ancient books in the digital age. [**Method/process**] The scriptures, biography and all index entries of *Index to Chunqiu Jingzhuan* was counted by digitization. Comparing them item by item was to find out and correct the errors and omissions, record and count the types and numbers of errors and omissions, and analyze the overall compilation quality. [**Result/conclusion**] There are few missing errors in *Index to Chunqiu Jingzhuan*, and the entry error rate is only about 1 in 10,000. The compilation quality can be called the pinnacle of the hand-crafted era, it deserves the great reputation among academic community and can be used as a high-quality foundation for digital corpus.

Keywords: *Index to Chunqiu Jingzhuan* revision compilation quality digitization manual corpus